Paper 047-31

SAS® Macros for Generating Abridged and Cause-Eliminated Life Tables

Zhao Yang, Xuezheng Sun

Department of Epidemiology & Biostatistics, University of South Carolina, Columbia, SC

ABSTRACT

The life table, a valuable tool in research, is a statistical table based on age-specific death rates in a particular population. The macros(`%abridge`, `%celiminate`) developed in this paper are used to generate abridged life tables and cause-eliminated life tables, respectively. These MACROS are easily implemented in practice.

Keywords: Abridged life table, Cause-eliminated life table, SAS macro

INTRODUCTION

The life table is a statistical table based on age-specific death rates in a particular population. For many years, life tables have been used as a means of summarizing the health status of a group of individuals. Life tables identify the death rates experienced by a population over a given period of time. They have many practical applications; life tables are used to analyze the mortality of a particular population, to make international comparisons, to compute insurance premiums and annuities, to predict survival, etc. Life tables can be categorized into abridged life tables and complete life tables according to different methods of classifying age. For abridged and cause-eliminated life table, the standard grouping for age has 19 groups: $0-1, 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, 60-65, 65-70, 70-75, 75-80, 80-85,$ and $85-$. The SAS MACRO developed in this paper will be based on this grouping method.

The cause-eliminated life table has almost the same table structure, indexes, and calculation method as the abridged life table. Because of limitations for cause-specific mortality rate(which can be influenced by age structure of the population), age-specific mortality rate(which can not provide a comprehensive evaluation to some cause on the mortality rate), and standardized mortality rate(which may not be compared for selecting different standard), the indexes derived from the cause-eliminated life table can provide a more precise and comprehensive evaluation of the cause on the expected lifetime. Therefore, since it is not affected by age-structure, it is better than the abridged life table for comparison and inference.

The idea behind generating life tables is to follow a fictional cohort of individuals—usually a group of $1,000,000$ persons—from birth until the last individual in the cohort has died. A life table describes the mortality experience of the group over a specified period of time.

CALCULATION METHODS

Our calculation method is based on the book written by Shuqing Yang, et al. Although there is a very small difference from other books, such as Marcello Pagano, in dealing with the calculations within the $0-1$ group, the resulting expected life time index is very close.

CALCULATION FOR ABRIDGED LIFE TABLE

The age-specific number in the population($_n\mathrm{P}_x$) and number of death($_n\mathrm{D}_x$) are required to generate an abridged life-table. Here, x indicates the starting point for an age interval, and n is the interval length$\big($e.g. $_5\mathrm{P}_{35}$ is for age interval $35-40\big)$. There are several important age-specific indexes in the abridged life table: mortality rate($_n\mathrm{m}_x$), mortality probability($_n\mathrm{q}_x$), expected survived population in each age-interval($_n\ell_x$),expected number of deaths in each age-interval($_n\mathrm{d}_x$), survived person-years($_n\mathrm{L}_x$), total survived person-year(T_x), and expected average remaining lifetime(e_x). The following are the formulas to calculate the above indexes.

$$_n\mathrm{m}_x = \frac{_n\mathrm{D}_x}{_n\mathrm{P}_x} \qquad _n\mathrm{q}_x = \frac{2 \times n \times {_n\mathrm{m}_x}}{2 + n \times {_n\mathrm{m}_x}} \qquad _n\mathrm{d}_x = {_n\ell_x} \times {_n\mathrm{q}_x}$$

$$\ell_{x+n} = {_n\ell_x} - {_n\mathrm{d}_x} \qquad \mathrm{L}_0 = \ell_1 + a_0 \times d_0 = \ell_1 + a_0 \times (\ell_0 - \ell_1) = (1 - a_0)\ell_1 + a_0 \times \ell_0 \qquad _n\mathrm{L}_x = \frac{n \times (\ell_x + \ell_{x+n})}{2}$$

$$\mathrm{L}_{80-} = \frac{\ell_{80}}{\mathrm{m}_{80-}} \qquad \mathrm{T}_x = \sum {_n\mathrm{L}_x} \qquad \mathrm{e}_x = \frac{\mathrm{T}_x}{\ell_x}$$

Generally, we do not need to calculate the mortality rate for 0− group, and for the mortality probability we can use the infant mortality rate or we can calculate it by $\mathrm{q}_0 = \frac{\mathrm{D}_0}{\mathrm{P}_0}$. Also the mortality probability for 80− is set as $\mathrm{q}_{80-} = 1.000000$, and $0 < a_0 < 1$ in the above formula is a coefficient, which varies among different countries. In our MACRO, we set it to 0.15.

CALCULATION FOR CAUSE-ELIMINATED LIFE TABLE

The calculation for cause-eliminated life table is almost the same as the abridged life table, but we have to adjust the cause in the calculation. Thus, there is some small difference between them. The age-specific number of population($_n\mathrm{P}_x$), number of death($_n\mathrm{D}_x$) and the number of death due to the cause($_n\mathrm{D}_x^i$) are required to generate a cause-eliminated life table. Here, i indicates some cause exists, while $(-i)$ means the cause is eliminated. There are several different important age-specific indexes in the cause-eliminated life table: the proportion of death due to other causes($_n\mathrm{r}_x$), survival probability after eliminating the cause($_n\mathrm{p}_x^{-i}$), expected survived population after eliminating the cause in each age-interval($_n\ell_x^{-i}$), survived person-years after eliminating the cause($_n\mathrm{L}_x^{-i}$), total survived person-year after eliminating the cause(T_x^{-i}), and expected average remaining lifetime after eliminating the cause(e_x^{-i}). The following are the formulas to calculate the above indexes.

$$_n\mathrm{r}_x = 1 - \frac{_n\mathrm{D}_x^i}{_n\mathrm{D}_x} \qquad _n\mathrm{p}_x^{-i} = \left({_n\mathrm{p}_x}\right)^{_n\mathrm{r}_x} = \left(1 - {_n\mathrm{q}_x}\right)^{_n\mathrm{r}_x} \qquad \ell_{x+n}^{-i} = {_n\ell_x^{-i}} \times {_n\mathrm{p}_x^{-i}}$$

$$\mathrm{L}_0^{-i} = \ell_1^{-i} + a_0 \times d_0^{-i} = (1 - a_0)\ell_1^{-i} + a_0 \times \ell_0^{-i} \qquad _n\mathrm{L}_x^{-i} = \frac{n \times (\ell_x^{-i} + \ell_{x+n}^{-i})}{2} \qquad \mathrm{L}_{80-}^{-i} = \frac{\ell_{80}^{-i}}{\mathrm{m}_{80-}^{-i}}$$

$$\mathrm{m}_{80-}^{-i} = \frac{\mathrm{D}_{80-} - \mathrm{D}_{80-}^{-i}}{\mathrm{P}_{80-}} \qquad \mathrm{T}_x^{-i} = \sum {_n\mathrm{L}_x^{-i}} \qquad \mathrm{e}_x^{-i} = \frac{\mathrm{T}_x^{-i}}{\ell_x^{-i}}$$

The mortality probability for 0− group is calculated by $\mathrm{q}_0 = \frac{\mathrm{D}_0}{\mathrm{P}_0}$. Also the mortality probability for 80− is set as $\mathrm{q}_{80-} = 1.000000$, and a_0 in the above formula is a coefficient, which varies among different countries. In our MACRO, we set it to 0.15.

DATA FOR LIFE TABLE GENERATION

We will use the following data(Shuqing, Yang, et al) to generate a related life table. It includes the age group, average population, real number of deaths, and the number of deaths due to the tumor for each group.

Age group	Average population	Real number of death	death due to tumor	Age group	Average population	Real number of death	death due to tumor
0-	30005	429	2	40-	90891	234	80
1-	86920	105	4	45-	105382	417	142
5-	102502	81	8	50-	86789	602	210
10-	151494	113	11	55-	69368	919	315
15-	182932	157	13	60-	51207	1328	360
20-	203107	215	21	65-	39112	1691	381
25-	190289	221	36	70-	20509	1561	248
30-	147076	181	41	75-	9301	1126	127
35-	99665	100	44	80-	4297	900	64
				85-	2714	1208	93

SAS MACRO

We will first illustrate the MACRO used to generate the abridged life table, to be followed by the one used to generate the cause-eliminated life table.

This MACRO involves some data step processing to generate the final data set for the life table, then via `PROC TEMPLATE` and `PROC REPORT` to create a standard statistical table for reporting.

MACRO—*%abridge* TO GENERATE ABRIDGED LIFE TABLE

```
%macro abridge(path, dsn, coeff, dlm, out);
proc format;
 picture agef
      low-high='09-';

data &dsn;
 infile "&path" dlm=&dlm;
 input age aver_p death;
      n+1;
           if n>1 then dr=death/aver_p;
           if age=1 then q=2*4*death/aver_p/(2+4*death/aver_p);
                else q=2*5*death/aver_p/(2+5*death/aver_p);
           if age=0 then q=death/aver_p;
           if age=85 then q=1.000000;
           p=1-q;
           if age=0 then sur=100000;

data temp1 ;
 set &dsn;
      retain x 1;
      x=x*p;
      xx=lag(x);
           if n=1 then xx=1;
           if n>1 then sur=100000*xx;

proc sort data=temp1;
 by descending n;

data temp2;
set temp1;
lagsur=lag(sur);
 if n=19 then lagsur=0.000000;
 death1=sur-lagsur;
```

The MACRO `%abridge` has several parameters: `path` indicates the data set path, since we read the data from an external file, like `.txt`; `dsn` indicates the data set name, which we want to create for the external file in SAS; `coeff`, as mentioned above, is a coefficient to calculate survived person-years for group 0−, which you can indicate this coefficient here. Considering the external files can have different delimiter values, we use `dlm` to indicate it; `out` is for the destination when the program runs the ODS, here, we create a `.pdf` file, which is easy to use for further work.

First, we use `PROC FORMAT` to create a format for each age group. Since we read the age group as a numeric variable, the value is the starting point of each interval, and we want to output the age variable as $0-, 5-$ etc. Then the followed program generates the `final` data set. It is only an implementation of the calculation method mentioned above. The only important thing to note here is the function `lag()`, which is necessary for calculations.

In the following two parts of the program, we then use `PROC TEMPLATE` to self-define our style. Since we want to create the standard statistical reporting table, this style is a modification of the build-in style template `styles.SansPrinter`. There are many standard styles for usage, such as `BarrettsBlue`, `FancyPrinter`, etc. For further information, you can refer to the useful paper by **Sunil K. Gupta** in SGUI26. Also, there is a good reference for using `PROC TEMPLATE` written by **Lauren Haworth**.

The last part of the program is to generate the final report by using `PROC REPORT`. We assume that you have some knowledge about this procedure, hence, the usage is pretty straightforward.

```
select(n);
     when(1) l=(1-&coeff)*lagsur+&coeff*sur;
     when(2) l=2*(sur+lagsur);
     when(19) l=sur/(death/aver_p);
     otherwise l=2.5*(sur+lagsur);
end;

data final;
 set temp2;
 t+l;
 e=t/sur;
proc sort data=final;
 by n; run;
proc template;
  define style self.border;
     parent=styles.SansPrinter;
       style Table /
         rules = groups
         frame=hsides
         cellpadding = 3pt
         cellspacing = 0pt
         borderwidth = 2pt;
     style header /
        font_weight=bold
        background=white
        font_size=3;
   end; run;

options nodate;
ods listing close;
ods pdf file="&out"  style=self.border;
```

```
title "Abridged life table for some district in 1982";
proc report data=final headline headskip nowd spacing=2 split='-' center;
   columns age aver_p death  q sur death1 l t e;
   define age    /display f=agef.
                   "Age--Group" width=5 center;
   define aver_p /display f=comma12.0
                   "Average--Population" width=10 center;
   define death  /display f=comma12.0
                   "Number-of-Death" width=9 center;
   define q      /display f= 12.6
                   "Death--Probability" width=11 center;
   define sur    /display f=comma12.0
                   "Number-of-Survived" width=11 center;
   define death1 /display f=comma12.0
                   "Expected-number-of Death" width=15 center;
   define l      /display f= comma12.0
                   "Survived-person-year" width=11 center;
   define t      /display f= comma12.0
                   "Total-Survived-person year" width=14 center;
   define e      /display f= 12.2
                   "Average-remaining-lifetime" width=18 center;
run;

ods pdf close;
ods listing;

%mend abridge;
%abridge(h:\raw.txt,life,0.15,'09'x, h:\life.pdf)
```

The following part is the output from the program. You can use the `Snapshot Tool` to make a copy in the output `life.pdf` file. It is easy! Then, we can make the output for data analysis report.
In our program, we only created a simple title in the above table, since our program is easy for repeated use and different data resource, you can easily add and modify the title in the `PROC REPORT` part.

MACRO—*%celiminate* TO GENERATE CAUSE-ELIMINATED LIFE TABLE

This macro is almost the same as the the `%abridge`, except for some small differences in the data step processing and in the `PROC REPORT` part, since the evaluation indexes in cause-eliminated life table are different than the abridged life table. We give macro `%celiminate` and its corresponding output in **Appendix-1**, and **Appendix-2**, respectively.

Finally, we checked our program with one different data set, from the book written by Marcello Pagano, et al. There is small difference between our results and those in the book, which can be attributed to slightly different calculation methods when dealing with the 0− group.

Abridged life table for some district in 1982

Age Group	Average Population	Number of Death	Death Probability	Number of Survived	Expected number of Death	Survived person year	Total Survived person year	Average remaining lifetime
0-	30,005	429	0.014298	100,000	1,430	98,785	6,888,545	68.89
1-	86,920	105	0.004820	98,570	475	393,331	6,789,760	68.88
5-	102,502	81	0.003943	98,095	387	489,508	6,396,430	65.21
10-	151,494	113	0.003723	97,708	364	487,632	5,906,921	60.45
15-	182,932	157	0.004282	97,345	417	485,681	5,419,289	55.67
20-	203,107	215	0.005279	96,928	512	483,359	4,933,609	50.90
25-	190,289	221	0.005790	96,416	558	480,685	4,450,249	46.16
30-	147,076	181	0.006134	95,858	588	477,819	3,969,565	41.41
35-	99,665	100	0.005004	95,270	477	475,157	3,491,746	36.65
40-	90,891	234	0.012790	94,793	1,212	470,934	3,016,589	31.82
45-	105,382	417	0.019591	93,581	1,833	463,319	2,545,655	27.20
50-	86,789	602	0.034091	91,747	3,128	450,917	2,082,336	22.70
55-	69,368	919	0.064117	88,619	5,682	428,892	1,631,419	18.41
60-	51,207	1,328	0.121775	82,937	10,100	389,438	1,202,527	14.50
65-	39,112	1,691	0.195088	72,838	14,210	328,664	813,089	11.16
70-	20,509	1,561	0.319726	58,628	18,745	246,278	484,424	8.26
75-	9,301	1,126	0.464675	39,883	18,533	153,084	238,146	5.97
80-	4,297	900	0.687338	21,350	14,675	70,065	85,062	3.98
85-	2,714	1,208	1.000000	6,675	6,675	14,998	14,998	2.25

CONCLUSION

The macros(`%abridge`, `%celiminate`) developed in this paper are used to generate abridged life tables and cause-eliminated life tables, respectively. They are developed to read an external data set file into SAS automatically.

Hence, for example, you can first input your data into `Microsoft Excel`, then save your data as a `Text(Tab Delimited)(*.txt)` file using the `Save as type` option(of course, you can save the data as another related readable file if you like, but you have to indicate a different `dlm` option in the macro). Then, you can run the macro to get the output report. At the same time, the macro is easy to modify for your convenience; it only provides a framework where your can self-define a life-table, if necessary.

References

[1] Lauren Haworth, *SAS with Style: Creating your own ODS Style Template for RTF Output*, SGUI29, Genentech, Inc., South San Francisco, CA

[2] Marcello Pagano, Kimberlee Gauvreau, *Principles of Biostatistics*, Duxbury: Thomson Learning, 2000, 2^{nd} edition, 98-104

[3] Shuqing Yang, el al, *Health Statistics*, Beijing: People Health Publishing Society, 1992, 3^{rd} edition

[4] Sunil K. Gupta, *Using Styles and Templates to Customize SAS ODS Output*, SGUI26, Gupta Programming, Simi Valley, CA

ACKNOWLEDGEMENTS

The authors are indebted to Dr. James R. Hussey for his valuable comments and suggestions on this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Zhao Yang
Department of Epidemiology & Biostatistics,
University of South Carolina
Columbia, SC 29208
803-777-8907
E-mail: `yang57@mailbox.sc.edu`

Xuezheng Sun
Department of Epidemiology & Biostatistics,
University of South Carolina
Columbia, SC 29208
803-777-8907
E-mail: `sun4@mailbox.sc.edu`

APPENDIX

Appendix-1: MACRO to generate cause-eliminated life table

```
%macro celiminate(path, dsn, coeff, dlm, out);
proc format;
    picture agef
        low-high='09-';

data &dsn;
    infile "&path" dlm=&dlm;
    input age aver_p death death1;
        n+1;
            if n>1 then dr=death/aver_p;
            pr=(death-death1)/death;
            if age=1 then q=2*4*death/aver_p/(2+4*death/aver_p);
                else q=2*5*death/aver_p/(2+5*death/aver_p);
            if age=0 then q=death/aver_p;
            if age=85 then q=1.000000;
            p=1-q;
            if n<19 then pi=p**pr;
                else  pi=0.000000;
            if age=0 then sur=100000;

data temp1 ;
    set &dsn;
        retain x 1;
        x=x*pi;
        xx=lag(x);
            if n=1 then xx=1;
            if n>1 then sur=100000*(xx);

proc sort data=temp1;
    by descending n;

data temp2;
set temp1;
    lagsur=lag(sur);
    if n=19 then lagsur=0.000000;
    death2=sur-lagsur;

    select(n);
        when(1) l=(1-&coeff)*lagsur+&coeff*sur;
        when(2) l=2*(sur+lagsur);
        when(19) l=sur/((death-death1)/aver_p);
        otherwise l=2.5*(sur+lagsur);
    end;

data final;
    set temp2;
    t+l;
    e=t/sur;

proc sort data=final;
    by n; run;
```

```
proc template;
  define style self.border;
     parent=styles.SansPrinter;
       style Table /
         rules = groups
         frame=hsides
         cellpadding = 3pt
         cellspacing = 0pt
         borderwidth = 2pt;
         style header /
         font_weight=bold
         background=white
         font_size=1.05;
   end; run;

options nodate;
ods listing close;

ods pdf file="&out" style=self.border;

title "Cause(tumor)-eliminated life table for some district in
1982";

proc report data=final headline headskip nowd spacing=2 split='-'
center;
    columns age aver_p death  death1 pr p
        ('__________Cause eliminated indexes_________' pi sur l t e);
    define age    /display f=agef.
                  "Age--Group" width=5 center;
    define aver_p /display f=comma12.0
                  "Average--Population" width=10 center;
    define death  /display f=comma12.0
                  "Number-of-Death" width=9 center;
    define death1  /display f=comma12.0
                  "Death due-to-tumor" width=9 center;
    define pr      /display f=comma12.6
                  "Death proportion-not-due to tumor" width=16 center;
    define p       /display f=comma12.6
                  "Total-survival-Probability" width=15 center;
    define pi      /display f= 12.6
                  "Survival--Probability" width=11 center;
    define sur    /display f=comma12.0
                  "Number-of-Survived" width=11 center;
    define l      /display f= comma12.0
                  "Survived-person-year" width=11 center;
    define t      /display f= comma12.0
                  "Total-Survived-person year" width=14 center;
    define e      /display f= 12.2
                  "Average-remaining-lifetime" width=18 center;
run;

ods pdf close;
ods listing;
%mend celiminate;
%celiminate(h:\raw1.txt,cause,0.15,'09'x, h:\cause.pdf)
```

Appendix-2: Cause-eliminated life table from MACRO-%celiminate

Cause(tumor)-eliminated life table for some district in 1982 1

						Cause eliminated indexes				
Age Group	Average Population	Number of Death	Death due to tumor	Death proportion not due to tumor	Total survival Probability	Survival Probability	Number of Survived	Survived person year	Total Survived person year	Average remaining lifetime
0-	30,005	429	2	0.995338	0.985702	0.985769	100,000	98,790	7,114,032	71.14
1-	86,920	105	4	0.961905	0.995180	0.995363	98,577	393,393	7,015,242	71.17
5-	102,502	81	8	0.901235	0.996057	0.996445	98,120	489,727	6,621,849	67.49
10-	151,494	113	11	0.902655	0.996277	0.996639	97,771	488,033	6,132,122	62.72
15-	182,932	157	13	0.917197	0.995718	0.996072	97,442	486,255	5,644,089	57.92
20-	203,107	215	21	0.902326	0.994721	0.995236	97,060	484,142	5,157,834	53.14
25-	190,289	221	36	0.837104	0.994210	0.995151	96,597	481,815	4,673,692	48.38
30-	147,076	181	41	0.773481	0.993866	0.995252	96,129	479,503	4,191,877	43.61
35-	99,665	100	44	0.560000	0.994996	0.997195	95,672	477,691	3,712,374	38.80
40-	90,891	234	80	0.658120	0.987210	0.991564	95,404	475,007	3,234,684	33.91
45-	105,382	417	142	0.659472	0.980409	0.987037	94,599	469,930	2,759,676	29.17
50-	86,789	602	210	0.651163	0.965909	0.977667	93,373	461,651	2,289,747	24.52
55-	69,368	919	315	0.657236	0.935883	0.957383	91,287	446,711	1,828,096	20.03
60-	51,207	1,328	360	0.728916	0.878225	0.909690	87,397	417,253	1,381,385	15.81
65-	39,112	1,691	381	0.774690	0.804912	0.845249	79,504	366,763	964,131	12.13
70-	20,509	1,561	248	0.841127	0.680274	0.723212	67,201	289,503	597,368	8.89
75-	9,301	1,126	127	0.887211	0.535325	0.574416	48,600	191,293	307,865	6.33
80-	4,297	900	64	0.928889	0.312662	0.339611	27,917	93,494	116,572	4.18
85-	2,714	1,208	93	0.923013	0.000000	0.000000	9,481	23,077	23,077	2.43